

PATENT ABSTRACTS OF JAPAN

(11)Publication number : 2000-148985

(43)Date of publication of application : 30.05.2000

(51)Int.Cl.

G06T 1/00
 A61B 5/117
 G06F 1/00
 G06F 3/00
 G06F 3/16
 G06F 15/00
 G06T 7/00
 G10L 17/00

(21)Application number : 10-326327

(71)Applicant : HITACHI LTD

(22)Date of filing : 17.11.1998

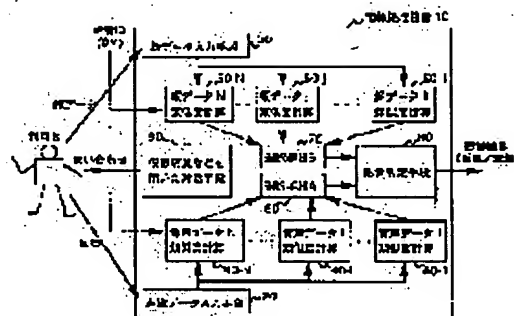
(72)Inventor : SATO YUJI
 OHIRA EIJI
 SUMINO SHIGEO

(54) PERSONAL AUTHENTICATION SYSTEM

(57)Abstract:

PROBLEM TO BE SOLVED: To acquire respective biological information data under the condition that is as much close to when data is recorded as possible by synthetically deciding from the final similarity to 1st biological information and the final similarity to 2nd biological information.

SOLUTION: An interacting means 90 performs several questions to a user. Then, a voice data inputting means 20 acquires voice data necessary to authentication from questions and answers with the user. Otherwise, a face data inputting means 30 acquires face data. A voice data similarity calculating means 40 calculates similarity by inputting a designated speaker ID and the voice data sent from the means 20 and sends the calculated similarity to a selecting means A 60. The means A 60 selects one among plural pieces of inputted similarities according to a preliminarily defined rule and sends it to a synthetically judging means 80. The means 80 performs synthetic judgement from the similarity of the voice data sent from the means A 60 and the similarity of face data sent from a selecting means B 70.



This Page Blank (uspto)

【特許請求の範囲】

【請求項 1】第 1 の生体情報を入力する手段と、第 1 の入力手段を介して得られた一つ以上の生体情報の類似度を求める手段と、上記一つ以上の第 1 の生体情報の類似度から最終的類似度を決定または選択する手段と、第 2 の生体情報を入力する手段と、第 2 の入力手段を介して得られた一つ以上の生体情報の類似度を求める手段と、上記一つ以上の第 2 の生体情報の類似度から最終的類似度を決定または選択する手段と、上記第 1 の生体情報に対する最終的類似度と第 2 の生体情報に対する最終的類似度から総合的な判定を行う手段と、利用者との対話手段を有することを特徴とする情報処理装置。

【請求項 2】第 1 の生体情報を入力する手段と、第 1 の入力手段を介して得られた一つ以上の生体情報の類似度を求める手段と、上記一つ以上の第 1 の生体情報の類似度から最終的類似度を決定または選択する手段と、第 2 の生体情報を入力する手段と、第 2 の入力手段を介して得られた一つ以上の生体情報の特徴量を求める手段と、上記一つ以上の第 2 の生体情報の特徴量と上記第 1 の生体情報に対する最終的類似度から総合的な判定を行う手段と、利用者との対話手段を有することを特徴とする情報処理装置。

【請求項 3】請求項 1 または請求項 2 の情報処理装置において、上記生体情報は、音声、眼、鼻、唇、耳、眉毛などの顔情報の一部または顔画像全体、筆跡、体臭のいずれかであることを特徴とする情報処理装置。

【請求項 4】請求項 1 または請求項 2 の情報処理装置において、上記一つ以上の生体情報の類似度から最終的類似度を決定する手段は、最大類似度を最終的な類似度とする手段であることを特徴とする情報処理装置。

【請求項 5】請求項 1 の情報処理装置において、上記一つ以上の生体情報の類似度から最終的類似度を決定する手段は、上位幾つかの類似度の平均値を最終的な類似度とする手段であることを特徴とする情報処理装置。

【請求項 6】請求項 1 の情報処理装置において、上記第 1 の生体情報に対する最終的類似度と第 2 の生体情報に対する最終的類似度から総合的な判定を行う手段は、学習機能を有することを特徴とする情報処理装置。

【請求項 7】請求項 2 の情報処理装置において、上記一つ以上の第 2 の生体情報の特徴量と上記第 1 の生体情報に対する最終的類似度から総合的な判定を行う手段は、学習機能を有することを特徴とする情報処理装置。

【請求項 8】請求項 6 の情報処理装置において、上記学習機能は、各生体情報に対応した重み値を各生体情報の最終的な類似度に掛けた値の組み合わせから総合的に判定する仕様とした上で、判定結果の正否に基づき、上記重み値を更新する機能であることを特徴とする情報処理装置。

【請求項 9】請求項 6 または請求項 7 の情報処理装置において、上記学習機能は、利用者ごとに設けることを特

徴とする情報処理装置。

【請求項 10】請求項 1 の情報処理装置において、上記第 1 の生体情報に対する最終的類似度と第 2 の生体情報に対する最終的類似度から総合的な判定を行う手段は、アクセス対象となる情報や応用ごとに判定の厳しさを可変とする機能を有することを特徴とする情報処理装置。

【請求項 11】請求項 2 の情報処理装置において、上記第 1 の生体情報に対する最終的類似度と第 2 の生体情報に対する一つ以上特徴量から総合的な判定を行う手段は、アクセス対象となる情報や応用ごとに判定の厳しさを可変とする機能を有することを特徴とする情報処理装置。

【請求項 12】請求項 1 または請求項 2 の情報処理装置において、上記利用者との対話手段は、仮想店員／仮想受付人などのエージェントを設け、利用者と同じエージェントとの対話の間に個人認証に必要な生体情報の入手を行う手段であることを特徴とする情報処理装置。

【請求項 13】請求項 1 2 の情報処理装置において、上記エージェントは、複数のキャラクタメニューの中から利用者が選択できる機能を有することを特徴とする情報処理装置。

【請求項 14】請求項 1 2 の情報処理装置において、上記エージェントの声は、録音した肉声またはテキストから機械的に合成した合成音声を用いることを特徴とする情報処理装置。

【請求項 15】請求項 1 4 の情報処理装置において、上記エージェントの声として日本語の合成音声を用いる場合には、ロボットや宇宙人などの自然な人間の声質から大きくずれていても違和感のないキャラクタを用いることを特徴とする情報処理装置。

【請求項 16】請求項 1 2 の情報処理装置において、上記エージェントの声は、利用者の好みに応じて自由に交換できる手段を有することを特徴とする情報処理装置。

【請求項 17】請求項 1 または請求項 2 の情報処理装置において、上記利用者との対話手段は、個人認証に使用する生体情報の種類をシステムに委任するか利用者自身が選ぶかの選択手段を有することを特徴とする情報処理装置。

【請求項 18】二つ以上の第 1 の生体情報データと二つ以上の第 2 の生体情報データを入力して個人認証を行う認証方式であって、異なる時刻に採取した第 1 の生体情報データと第 2 の生体情報データの組み合わせから個人認証を行うことを特徴とする個人認証方式。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】本発明は、情報処理装置のセキュリティ向上方法に関する。より詳細には、テレホンバンキング、インターネットバンキングやインターネットショッピングなどのための情報処理装置における本人認証精度の向上およびそのためのユーザインタフェー

スの実現方法とその他の装置に関する。

【0002】

【従来の技術】銀行の預金引き落としや商品の購入などにおいて、特定の個人を認証するための手段として、磁気ストライプを有するカード（以下、磁気カードと呼ぶ）が広く用いられている。キャッシュカードにより銀行の自動預け払い機から現金を引き出す場合、暗証番号により本人の特定を行う。情報処理装置内部に格納された情報の機密保持手段としては、情報へのアクセスに要求者に対して、パスワードなどの入力でアクセス権限の確認を行っている。

【0003】一方、上記個人認証方法に対してセキュリティに関する問題点が指摘されている。例えば、暗証番号やパスワードが第三者に覚えられたり解読されたりした場合、預金が不正に引き出される、あるいは、機密情報が盗まれる恐れがある。この問題を恐れて複雑な暗証番号やパスワードにすると、忘却や入力ミスなどの恐れがある。

【0004】そこで、暗証番号やパスワードに代えて、指紋などの生体情報を用いた個人認証方法およびそのための製品が普及し始めている。生体情報は第三者に盗まれる恐れが低く、かつ忘れる心配もないためである。

【0005】しかし、現在最も製品化の進んでいる指紋は、指紋の登録が犯罪者を連想させるなど、必ずしも利用者にとって心象が良くない。また、認証に際して、「これから認証される」という印象を強く利用者を与えてしまう。指紋以外で認証精度の高い生体情報としては、網膜パターンや掌形などもあるが、読み取り装置が、数十万円～数百万円と高価になる。

【0006】利用者に受け入れられ易く、かつ装置が比較的安価となる生体情報として声紋、顔画像、サインなどが存在するが、これらの生体情報を用いた個人認証では、例えば、データ登録時と認証時との顔とマイク（またはテレビカメラ）との微妙な角度や距離の差から、認証精度が実用上十分にならないという問題があった。

【0007】そこで、顔画像と音声の両者から個人認証を行うアイデアが、磯部俊洋、坂野鋭、劉偉傑：「顔画像と音声を用いた個人認証方式」、日本音響学会講演論文集、pp. 71-72 (1998. 3) として提案されている。

【0008】このアイデアを説明するために、まず一般的な話者認識方式について簡単に説明する。一般的な話者認識システムに関しては、例えば、古井著：「音響・音声工学」、近代科学社（1992）のpp. 211-241に詳しく述べられている。話者認識の形態は、話者識別と話者照合に分けることができる。

【0009】図2に話者識別システムの構成例を示す。図2において、話者識別システム200は、特徴抽出部210、類似度計算部220-1～220-N、話者ごとの標準パターン230-1～230-N、最大値選択部240から構成されている。話者識別システムとは、入力音声が予め登録されて

いるN人のうちの誰の声であるかを判定するシステムである。まず、特徴抽出部210により、入力音声の特徴量データが抽出される。この特徴量データと予め登録している話者N人分の標準パターン230-1～230-Nとの比較により、登録話者それぞれに対する類似度が類似度計算部220-1～220-Nにより求められる。最大値選択部240では、最も類似度の高い話者が、その話者の音声であると判断して認識結果を、例えば話者番号として、出力する。

【0010】図3に話者識別システムの構成例を示す。図3において、話者識別システム100は、特徴抽出部110、類似度計算部120、標準パターン130、判定部140、しきい値150から構成されている。話者照合システムとは、入力音声と同時に自分が誰であるかのIDを入力して、その音声の確かにそのIDに対応する人の音声であるか否かを判定するシステムである。まず、特徴抽出部110により、入力音声の特徴量データが抽出される。この特徴量データと、話者識別システム100の外部から入力された話者IDに対応した、予め登録している標準パターン130との比較を類似度計算部120で行い、類似度を求める。判定部140では、この類似度がしきい値150よりも大きければ本人の音声であると判定し、それ以外の場合は他人の音声であると判断して認識結果、例えば受理／棄却の区別、を出力する。

【0011】しきい値150の設定方法は幾つか提案されている。例えば、本人が棄却される誤り率と他人が受理される誤り率から設定される。これら2種類の誤り率と判定のしきい値150との関係の例を図4に示す。図4に示すように、一般的には、しきい値150を大きくすれば他人が受理される誤り率が大きくなる。逆に、しきい値150を小さくすれば本人が棄却される誤り率が大きくなる。両者の誤りにはトレードオフの関係があるために、一般的には、本人棄却率と他人受理率が等しくなる値をしきい値150として設定する。あるいは、2種類の誤りの相対的な重要性に従って設定される。例えば、本人棄却率が多少大きくなっても、他人受理率がある一定値を越えないように設定される。

【0012】上記話者認識システムは利用者が気軽に利用でき、かつ装置が安価に実現できるという利点を持つ。一方、経年変化、背景雑音、マイクの種類、顔の向き、風邪による声質の変化などにより、性能が劣化して十分な認証精度を得ることが困難である。この問題を解決するための一つのアイデアが、前記文献「顔画像と音声を用いた個人認証方式」の技術である。話者認識におけるこのような問題と同様に、画像を用いた個人認証においても、照明条件、顔の向き、表情の変動により、性能が劣化して十分な認証精度を得ることが困難という問題が存在するが、音声認証器と顔画像認証器の劣化要因が必ずしも互いに相関を持たないと考え、この2種類の認証方式の統合により個人認証精度の向上を図るアイデ

アである。

【0013】このアイデアによる総合判定手段を図5に示す。図5に示す総合判定手段80は、顔画像用しきい値81、音声用しきい値82、2次元スコア空間での判定83から構成される。2次元スコア空間での判定83は顔の類似度 f 、音声の類似度 v 、顔画像用しきい値 $Tf81$ 、音声用しきい値 $TV82$ から受理／棄却を決定して認識結果として出力する。

【0014】図6に、顔・音声の照合スコアに基づく2次元スコア空間の例を示す。図6において、領域Aは顔画像、音声ともに類似度がしきい値を越えている領域であり、判定結果は受理となる。領域Dは、顔画像、音声ともに類似度がしきい値を下回っており、判定結果は棄却となる。領域Bは、顔画像の類似度はしきい値を越えているが、音声の類似度がしきい値を下回っている領域。領域Cは、音声の類似度がしきい値を越えているが、顔画像の類似度がしきい値を下回っている領域である。領域B、Cの判定に関しては、誤って他人を受理することを避けることを優先して、棄却とする仕様が考えられる。また、識別境界の設定を行い、受理／棄却の判定を行う提案もされている。

【0015】図7を用いて、従来例における、図6の領域Bでの識別境界の設定方法を説明する。図7において、領域Bに属する学習サンプルのうち、本来受理されるべき本人のカテゴリに属するものは△の記号で示している。▲は△の重心を表す。一方、本来棄却されるべき他人のカテゴリに属するものは○で表している。●は○の重心を表す。ここでは、重心▲を受理用テンプレート、重心●を棄却用テンプレートとする。認証時には、入力サンプルとユークリッド距離が小さいテンプレートのインデックスを判定結果とする。図6の領域Cに関しても同様の処理を行う。

【0016】

【発明が解決しようとする課題】上記従来技術では、音声と画像のそれぞれの単体精度がある程度高い場合には、音声単体、顔画像単体で認証を行う場合に比べて、認証精度の向上が図れる傾向を有している。一方、音声と画像のうち少なくともどちらか一方の精度が十分得られないと、音声単体、顔画像単体で認証を行う場合に比べて、逆に認証精度が低下する恐れを有している。従来例における実験報告でも、顔画像単体の誤受理率よりも、提案手法による誤受理率の方が高くなっている。顔と音声の両方同時に一回の試行で十分な条件でデータが取得できるとは限らないため、音声単体、顔画像単体で認証を行う場合に比べて、必ずしも精度が向上するとは限らないという問題が存在する。

【0017】また、図7に示した識別境界の設定は予め学習用サンプルを用いた設定を前提としている。しかし、実運用に十分な学習サンプルが利用者全員に対して予め十分得られる可能性は極めて低い。統計上十分な学

習サンプルが得られないまま、あるいは、精度の低い学習サンプルが含まれたまま識別境界の設定を行うと、実用上十分な識別境界の精度が得られないという問題点を有する。

【0018】本発明の目的は、これら上記従来技術の問題点を対策することにある。すなわち、第1の目的は、データ登録時とできるだけ近い条件下でそれぞれの生体情報データの取得を可能とすることにある。第2の目的は、実運用に十分な学習サンプルが予め十分得られなかったとしても、運用中に修正できる機能を提供することにある。さらに、第3の目的は、「これから認証される」という印象を強く利用者に与えることなく上記2つの目的を実現するためのユーザインタフェース機能を提供することにある。

【0019】

【課題を解決するための手段】上記問題点を解決するために、以下に示す手段を設ける。

【0020】第1に、一回の個人認証に最低限必要なデータを採取するための時間よりも長時間データを採取し続けて、個人認証を継続的に行う手段を設ける。例えば、複数回の個人認証分の生体情報データを記憶する手段、あるいは、長時間の時系列情報を記憶する手段を設ける。また、これら複数の生体情報の特徴量を抽出して予め登録した標準パターンとの間の類似度を求めた後、これら複数の類似度の中から一つを選択する手段を設ける。

【0021】第2に、複数種類の生体情報の類似度から総合的に判定処理を行う手段に学習機能を追加する。例えば、生体情報の信頼性に優位差がある場合、信頼性に対応した重みを各生体情報の類似度に掛けた値の組み合わせから総合的に判定する仕様とした上で、判定結果の正否を基にこの重み値を更新する機能を設ける。あるいは、各生体情報の類似度から総合的な判定を行うための関数を考え、判定結果の正否を基にこの関数の係数を更新する機能を設ける。また、これら重み値などの係数は利用者ごとに可変とする。

【0022】第3に、利用に際して比較的ユーザが違和感のない情報あるいは意識しないでもデータが採取可能な情報、例えば、音声、顔全体、眼、鼻、唇、耳、眉毛などの顔画像、署名、体臭などの生体情報を前提とする。また、仮想店員／仮想受付人などのエージェントを設け、利用者と仮想店員との対話の間に個人認証に必要な情報の入手を行う手段を設ける。例えば、パソコンを用いてインターネットに接続し上記買い物を行う場合、内蔵マイクを用いてユーザの音声データを取得して認証に用いる。あるいは、パソコンに取り付けたカメラで顔情報をとり続けて認証に用いる。パソコンの内蔵マイクとカメラが仮想店員の耳と眼を代行する仕様とする。

【0023】本発明では、第1に、一回の個人認証に最低限必要なデータを採取するための時間よりも長時間デ

ータを採取し続けて、個人認証を継続的に行う手段を有する。例えば、複数回の個人認証分の生体情報データを記憶する手段を有する。従って、一回のデータだけで個人認証をする従来例と比較して、複数回分のデータの中に、データ登録時と近い条件で取得したデータが含まれる可能性が高くなる。あるいは、長時間の時系列情報を記憶する手段を有する。従って、例えば顔画像データなど、1枚の2次元静止画像を基にした場合に考えられる、特徴量抽出失敗の不完全データによる照合を回避できる。また、これら複数の生体情報の特徴量を抽出して予め登録した標準パターンとの間の類似度を求めた後、これら複数の類似度の中から一つを選択する手段を有する。すなわち、異なる時刻の複数生体情報の組み合わせが可能となる。例えば、時刻*i*の音声による認証と時刻*j*の顔画像による認証の組み合わせが可能となるために、同一時刻の2つの情報の組み合わせを考える従来例よりも、精度の高いものどうしの組み合わせが可能となる。

【0024】第2に、複数種類の生体情報の類似度から総合的に判定処理を行う手段に学習機能を有する。例えば、生体情報の信頼性に優位差がある場合、信頼性に対応した重みを各生体情報の類似度に掛けた値の組み合わせから総合的に判定する仕様とした上で、判定結果の正否を基にこの重み値を更新する機能を有する。あるいは、各生体情報の類似度から総合的な判定を行うための関数を考え、判定結果の正否を基にこの関数の係数を更新する機能を有する。従って、統計上十分な学習サンプルが得られないまま識別境界の設定を行った場合でも、運用中に、より精度の高い識別境界に修正していくことができる。また、これら重み値などの係数は利用者ごとに可変とすることができ、よりきめ細かい制御が可能となる。

【0025】第3に、利用に際して比較的ユーザが違和感のない情報あるいは意識しないでもデータが採取可能な情報、例えば、音声、顔全体、眼、鼻、唇、耳、眉毛などの顔画像、署名、体臭などの生体情報を前提とする。また、仮想店員／仮想受付人などのエージェントを設け、利用者と仮想店員との対話の間に個人認証に必要な情報の入手を行う。例えば、パソコンを用いてインターネットに接続し上記買い物を行う場合、内蔵マイクを用いてユーザの音声データを取得して認証に用いる。あるいは、パソコンに取り付けたカメラで顔情報をとり続けて認証に用いる。従って、「これから認証される」という印象を強く利用者を与えることなく個人認証に必要な生体情報の取得が可能である。

【0026】

【発明の実施の形態】本発明の第1の実施例となる情報処理装置10の構成を図1に示す。情報処理装置10は、マイクなどの音声データ入力手段20、テレビカメラなどの顔データ入力手段30、音声データ類似度計算40-1~40-N、顔データ類似度計算50-1~50-N、選択手段A6

0、選択手段B70、総合判定手段80、対話手段90から構成されている。対話手段90は、仮想店員などのエージェントを用いて、利用者に対して製品紹介などを行う。あるいは、利用者へいくつかの質問を行う。対話手段90と利用者との間の質疑応答から、認証に必要な音声データを音声データ入力手段20により取得する。あるいは、顔データ入力手段30により顔データを取得する。音声データ類似度計算40は、指定された話者IDと音声データ入力手段20から送られてくる音声データを入力として類似度を求め、求めた類似度を選択手段A60に送る。選択手段A60では、予め決められた規則に従い、入力された複数の類似度の中から一つを選択して総合判定手段80に送る。総合判定手段80では、選択手段A60から送られてきた音声データの類似度と選択手段B70から送られてきた顔データの類似度から総合的な判定を行い、受理／棄却の判定結果を出力する。顔データ類似度計算50、選択手段B70に関しては、基本的な機能は音声の場合と同様であるため、ここでは説明を省略する。

【0027】類似度計算システム40の構成例を図8に示す。図8において、類似度計算システム40は、特徴抽出部110、類似度計算部120、標準パターン130から構成されている。まず、特徴抽出部110により、声紋データ入力手段20から送られてくる入力音声の特徴量データが抽出されて類似度計算部120に送られる。この特徴量データと、情報処理装置10の外部から類似度計算システム40に入力された話者IDに対応した、予め登録している標準パターン130との比較を類似度計算部120で行い、類似度を求める。この類似度は選択手段A60に送られる。

【0028】図9に最大類似度を選択する場合の選択手段Aの一例を示す。図9に示す選択手段A60は、音声類似度記憶手段61-1~61-Nおよび最大値検出手段62から構成されている。音声類似度記憶手段61-1~61-Nは、それぞれ、類似度計算システム40-1~40-Nから送られてくる類似度を記憶するメモリである。音声類似度記憶手段61-1~61-Nに記憶された類似度はそれぞれ最大値検出手段62に送られ、最大値検出手段62で最大値が音声類似度として選ばれて総合判定手段80に送られる。最大値を求めるアルゴリズムは幾つか存在する。例えば、61-1から送られてきた類似度をまず初めに仮の最大値とする。次に、61-2から送られてきた類似度と仮の最大値を比較し、仮の最大値の方が小さければ、61-2から送られてきた類似度を仮の最大値として入れ替える。同様の処理を順次61-Nまでくり返し、最後に仮の最大値として記憶された類似度が最終的な最大値となる。

【0029】図10に上位*k*個の類似度の平均値を出力として選択する場合の選択手段Aの一例を示す。図10に示す選択手段A60は、音声類似度記憶手段61-1~61-N、ソート手段63および平均値計算手段64から構成されている。音声類似度記憶手段61-1~61-Nは、それぞれ、類似度計算システム40-1~40-Nから送られてくる

類似度を記憶するメモリである。音声類似度記憶手段61-1~61-Nに記憶された類似度はそれぞれソート手段63に送られ、類似度の高い順番に並び替えられた後、類似度の高い上位k個が平均値計算手段64に送られる。平均値計算手段64では、ソート手段63から送られてくるk個の類似度の平均を求め、その値を音声類似度として総合判定手段80に送る。ソート方法に関しては、クイックソート、バブルソート等、幾つか提案されている汎用のソートアルゴリズムのいずれかを用いる。

【0030】図9、図10では、N回分の入力データに対応したN個の類似度データの中なら最終的な類似度データを求めて判定する方法の例を示した。N回分のデータ入力から得られるN個の特徴量を時系列データとして扱い、ニューラルネットワーク技術などを用いて、判定を行うこともできる。例えば、顔画像を対象とした場合の例が、玉井洋行、永野俊：「顔の時系列画像を用いた個人照合ニューラルシステム」、信学技報、NC97-162, pp. 179-184 (1998)に記載されている。一般に、顔を用いた個人照合の多くは1枚の2次元静止画像の正面顔を入力データとして特徴量の抽出を行い類似度計算を行う。一方、人間の顔には目や口といった部分に明確な特徴があるにも関わらず、撮影時の姿勢や表情、照明条件により見え方は様々に変化する。従って、1枚の2次元静止画像の正面顔を入力データとした場合、撮影時の姿勢や表情、照明条件の変動の影響を大きく受けてしまう。そこで、撮影時の姿勢や表情、照明条件の変動の影響を回避し、特徴抽出失敗の不完全データによる照合防止のために上記時系列データによる顔照合の提案がされている。また、多少の変動にも柔軟に対応できるようにニューラルネットワーク技術によるロバスト性の向上を図っている。

【0031】顔の時系列画像と音声データから個人照合を行う、第2の実施例となる情報処理装置11の構成を図19に示す。情報処理装置11は、マイクなどの音声データ入力手段20、テレビカメラなどの顔データ入力手段30、音声データ類似度計算40-1~40-N、顔データ特徴抽出51-1~51-N、選択手段A60、総合判定手段81、対

$$T_0 = \alpha T_v + (1 - \alpha) T_f, \quad (0 < \alpha < 1) \quad \dots\dots (1)$$

また、認証時の音声類似度を v 、顔類似度を f 、重み値を α として、以下に示す数2で定義される総合類似度を T を定義する。

$$T = \alpha v + (1 - \alpha) f, \quad (0 < \alpha < 1) \quad \dots\dots (2)$$

T が T_0 よりも大きければ受理と判定し、小さければ棄却と判定する。例えば、音声類似度 v が音声のしきい値 T_v よりも大きく、かつ顔類似度 f が顔のしきい値 T_f よりも大きい場合には、 α の値に依らず T は T_0 よりも大きくなり、判定結果は受理となる。 v が T_v よりも小さく、かつ f が T_f よりも小さい場合には、 α の値に依らず T は T_0 よりも小さくなり、判定結果は棄却となる。 v が T_v よりも大きく、かつ f が T_f よりも小さい場合、あるいは、 v が T_v よ

りも小さく、かつ f が T_f よりも大きい場合には、判定結果は α の値に依存する。 α を大きくする程、音声認証の影響を強く受けることになるため、話者照合の判定を誤ったことが原因で総合判定を誤った場合には α の値を微小量、例えば 0.01α だけ小さくした値を新たに α とする更新処理を行う。顔認識の判定を誤ったことが原因で総合判定を誤った場合には α の値を、例えば 0.01α だけ、大きくした値を新たに α とする更新処理を行う。

【0032】上記本発明によれば、選択手段A60、B70から総合判定手段80に送られる特徴量は従来例よりもデータ登録時と近い条件で採取された入力データに対応する確率が高くなる。従って、図6で示した領域B、Cに入る確率を低下させ、逆に、領域A、Dに入る確率を増加させることができる。

【0033】次に、総合判定手段80の構成例について示す。第1の構成例としては、図5、図6、図7で紹介した方式をそのまま流用することが考えられる。上記本発明によれば、選択手段A60、B70から総合判定手段80に送られる特徴量は従来例よりもデータ登録時と近い条件で採取された入力データに対応する確率が高いために、図6で示した領域B、Cに入る確率が低下する。すなわち、図5、図6、図7で紹介した方式をそのまま流用したとしても、認識結果の精度は向上する。

【0034】より認識結果の精度を向上させるためには、複数種類の生体情報の類似度から総合的に判定処理を行う総合判定手段80に学習機能を設けることが考えられる。例えば、生体情報の信頼性に優位差がある場合、信頼性に対応した重み値を各生体情報の類似度に掛けたいの組み合わせから総合的に判定する仕様とした上で、判定結果の正否を基にこの重み値を更新する機能を設ける。例えば、音声のしきい値を T_v 、顔のしきい値を T_f 、重み値を α とした時、以下に示す数1で定義される総合しきい値 T_0 を定義する。

【0035】

【数1】

【0036】

【数2】

【0037】これは、例えば図6の領域Bについて述べると、図11に示すように領域Bを2分する直線を考え、この直線より類似度が上に位置すれば受理、下に位置すれば棄却と判定する仕様とし、判定結果を誤った場合、この直線を修正することに対応する。 α を大きくすると、縦軸との交点を上方向にシフトし、横軸との交点を左方向にシフトすることに対応する。

【0038】この場合の総合判定手段80Bの一構成例を図12に示す。音声および顔の類似度から照合判定を行う方法に関しては、識別境界の設定方法が図7に代わって図11になった以外は図5と同じであるために説明を省略する。図5との違いは重み値 α の学習機能を実現するために、 α の記憶手段84と更新手段85を有する点にある。更新手段85は、総合判定手段80Bに対して入力された誤り通知信号と2次元スコア空間83から入力した話者および顔に関する判定結果をもとに、どちらの判定が誤っていたのかを解析し、記憶手段84に格納された α の増減を行う。 α の初期値としては、2種類の生体情報の影響を等しく受けるように、例えば0.5に設定する。記憶手段84は、例えば、情報処理装置内のメモリを用いる。ここでは、音声と顔画像の2種類の組み合わせに関して一例を示した。他の生体情報の組み合わせ、あるいは、3種類以上の生体情報の組み合わせに関しても同様の考え方で対応できる。

【0039】識別境界を決定する判別式に関しては必ずしも直線である必要はない。図13に示す様な曲線あるいは多項式でもよい。N種類の生体情報を対象としたN変数の多項式の場合は上記の場合と同様にN個の係数を学習することにより実現できる。あるいは、ニューラルネットワークを用いて係数の学習を行うことも可能である。

【0040】人により、顔の特徴が出にくい場合や声の特徴が出にくい場合がある。あるいは、適切なしきい値や識別境界関数が変わってくる可能性がある。そこで、図14に示すように、話者ごとに、顔しきい値、声しきい値、重み値、識別境界を決定する判別式を設けることも考えられる。

【0041】図14において、総合判定手段80Cは、話者ごとに、顔しきい値、声しきい値、重み値、識別境界を決定する判別式を記憶するメモリ86、2次元スコア空間での判定83C、更新手段85Cから構成される。2次元スコア空間での判定83Cは顔の類似度 f 、音声の類似度 v 、話者IDに対応した顔しきい値 T_f 、声しきい値 T_v 、重み値 α 、識別境界を決定する判別式から受理／棄却を決定して認識結果として出力する。更新手段85Cは、総合判定手段80Cに対して入力された誤り通知信号と2次元スコア空間83Cから入力した話者および顔に関する判定結果をもとに、どちらの判定が誤っていたのかを解析し、メモリ86に格納された、指定された話者IDに対応した変数の更新を行う。

【0042】上記本発明によれば、複数種類の生体情報の類似度から総合的に判定処理を行う手段に学習機能を有する。従って、統計上十分な学習サンプルが得られないまま上記領域B、Cの識別境界の設定を行った場合でも、運用中に、より精度の高い識別境界に修正していくことができる。また、識別境界の設定は利用者ごとに可変とすることができるために、よりきめ細かい制御が可能となる。

【0043】図15に、アクセス対象となる情報あるいは応用ごとに判定の厳しさを可変とする場合の総合判定手段80Dの構成例を示す。図15において、総合判定手段80Dは、応用ごとに、顔しきい値、声しきい値、重み値、識別境界を決定する判別式を記憶するメモリ87、2次元スコア空間での判定83D、から構成される。2次元スコア空間での判定83Dは顔の類似度 f 、音声の類似度 v 、応用IDに対応した顔しきい値 T_f 、声しきい値 T_v 、重み値 α 、識別境界を決定する判別式から受理／棄却を決定して認識結果として出力する。応用IDは、例えば利用者がどの応用、あるいは、どの情報にアクセスしようとしているか検知可能な対話手段90から得ることが可能である。

【0044】上記発明によれば、比較的機密性の低い情報はしきい値を下げて、本人が棄却する確率を減少させることができる。一方、極秘情報はしきい値を上げて、他人を誤って受理する確率を下げるができる。

【0045】図16に、仮想店員／仮想受付人などのエージェントを設け、利用者と仮想受付人との対話の間に個人認証に必要な情報の入手を行う例を示す。図16の例では、インターネットバンキングで振り込み処理を行う例を示している。生体情報としては音声と顔画像を仮定している。パソコンなどの情報処理装置の画面上には、仮想受付人が表示されている。仮想受付人は、振り込み処理に必要な情報の問い合わせ910を利用者に対して行う。利用者は、仮想受付人の問い合わせに対して応答920を音声で行う。他人に聞かれたくない内容は、一部、テキスト入力する。利用者の回答の内、マイクなどを通して音声入力された内容が話者照合用のデータとして適用される。例えば、「振り込みです」、「日立太郎です」、「東海銀行です」などの音声情報をもとに、それぞれ類似度計算が行われる。仮想受付人と利用者の対話の間に、並行してテレビカメラなどで顔画像を取り込み、顔画像による類似度を求める。図16において、仮想受付人からの問い合わせ内容は、仮に実際に銀行に向いて振り込み処理を行う場合でも通常必要な情報であるために、利用者に対して特に認証されているという強い印象を与えることがない。

【0046】図17に、インターネットショッピングで検索を行う例を示す。図16と同様に、生体情報としては音声と顔画像を仮定している。パソコンなどの情報処理装置の画面上には、仮想店員が表示されている。仮想

店員は、サービスを受けることができる登録会員であるかどうかの確認および利用者が知りたい内容の問い合わせ911を利用者に対して行う。利用者は、仮想店員の問い合わせに対して応答921を音声で行う。他人に聞かれない内容は、一部、テキスト入力する。利用者の回答の内、マイクなどを通して音声入力された内容が話者照合用のデータとして適用される。例えば、「はい」、「日立太郎です」、「製品紹介です」などの音声情報をもとに、それぞれ類似度計算が行われる。仮想店員と利用者の対話の間に、並行してテレビカメラなどで顔画像を取り込み、顔画像による類似度を求める。

【0047】上記図16、図17において、仮想店員の声は、肉声を録音して再生することで実現できる。あるいは、機械的に合成した合成音声を用いることもできる。仮想店員／仮想受付人は、予め録画した人間の映像、コンピュータグラフィック、人形、アニメなどで実現することができる。幾つかのキャラクタのメニューを準備し、利用者が選択する仕様とすることもできる。コンピュータ テレフォニーインテグレーション (Computer Telephony Integration) 技術を用いてパソコンを用いたテレビ電話を使えば、リアルタイムで実際の店員／受付人が対応することもできる。仮想店員の声として機械的に合成した日本語合成音声を用いる場合には、仮想店員のキャラクタとしてロボットまたは宇宙人などの自然な人間の声質から大きくずれていても違和感のないキャラクタを用いることもできる。また、特開平10-97267号公報「声質変換方法および装置」の技術を用いれば、予め登録された仮想店員／受付人の声質を、利用者の好みに応じて自由に交換ができる仕様とすることも可能である。

【0048】上記図16、図17の例においては、生体情報として音声と顔画像を仮定したが、利用に際して比較的ユーザが違和感のない情報あるいは意識しなくてもデータを採取可能な情報、例えば、音声、顔全体、眼、鼻、唇、耳、眉毛などの顔画像、署名、体臭などの生体情報は全て本発明の対象とすることができる。

【0049】上記本発明によれば、利用に際して比較的ユーザが違和感のない情報あるいは意識しなくてもデータを採取可能な情報、例えば、音声、顔全体、眼、鼻、唇、耳、眉毛などの顔画像、署名、体臭などの生体情報を前提とする。また、仮想店員／仮想受付人などのエージェントを設け、利用者と仮想店員との対話の間に個人認証に必要な情報を入手する。従って、「これから認証される」という印象を利用者に与えることなく個人認証に必要な生体情報の取得が可能である。

【0050】上記実施例では、利用者に対して出来るだけ気付かれないように個人認証を行う場合の例を示した。個人認証を全てシステムに任せるか、個人認証に使用する生体情報の種類を利用者自身が指定するかを、利用者に明示的に選択させる仕様とすることもできる。こ

の場合の処理の流れの一例を図18に示す。

【0051】図18の例では、まず、個人認証に使用する生体情報の種類をシステムに委任するか、利用者自身が決めるかの選択1000をシステムから利用者に問い合わせる。システムに委任1010を利用者が選択した場合は、音声、顔画像、体臭などによるシステム側主導の個人認証が行われる。利用者自身が決める1020を利用者が選択した場合は、選択可能な生体情報のメニューが表示される。この場合、利用者は、例えば指紋、音声、顔画像、サインの中から一つ以上を選択する1030。指紋が選択された場合は指紋による個人認証1031が実行される。音声を選択された場合は音声による個人認証1032が実行される。顔画像が選択された場合は顔画像による個人認証1033が実行される。サインが選択された場合はサインによる個人認証1034が実行される。上記処理は、判定結果の確信度が十分になるまで繰り返され1040、最後に受理／棄却の通知1050が出力される。

【0052】図18では一種類の生体情報を利用者が指定する場合の例を示したが、当然、二種類以上の生体情報を利用者が指定する仕様とすることもできる。また、使用対象から除外する生体情報を利用者に指定させる仕様とすることもできる。

【0053】個人認証に用いる生体情報を利用者が選択できる機能を設けることにより、例えば、風邪で声質に変化がある場合は声紋以外を認証に使用する、指に怪我をした場合はサイン以外を認証に使用する、女性など化粧の出来映えが顔データ登録時と大きく異なる場合は顔データ以外を使用するなど、認証精度の低下を防ぐ効果が期待できる。

【0054】

【発明の効果】本発明によれば、第1に、データ登録時とできるだけ近い条件下でそれぞれの生体情報データの取得を可能とすることができる。また、それらの最適な組み合わせができる。第2に、実運用に十分な学習サンプルが予め十分得られなかったとしても、より精度が良くなるように、運用中にパラメータの修正が可能である。第3に、「これから認証される」という印象を強く利用者に与えることなく上記2つの目的を実現するための生体情報を取得することができる。

【図面の簡単な説明】

【図1】本発明の第1の実施例となる情報処理装置を示すブロック図。

【図2】話者識別システムの機能構成例を示すブロック図。

【図3】話者照合システムの機能構成例を示すブロック図。

【図4】話者照合の判定のしきい値と2種の誤り率の関係を示す説明図。

【図5】総合判定手段の機能構成例を示すブロック図。

【図6】顔・音声の照合スコアに基づく2次元のスコア

空間の説明図。

【図7】図6の領域Bにおける識別境界の設定例の説明図。

【図8】類似度計算システムの機能構成例を示すブロック図。

【図9】選択手段の第1の機能構成例を示すブロック図。

【図10】選択手段の第2の機能構成例を示すブロック図。

【図11】図6の領域Bにおける識別境界の第2の設定例の説明図。

【図12】重み値学習機能付き総合判定手段の機能構成例を示すブロック図。

【図13】図6の領域Bにおける識別境界の第3の設定例の説明図。

【図14】変数学習機能付き総合判定手段の機能構成例を示すブロック図。

【図15】応用依存型総合判定手段の機能構成例を示すブロック図。

【図16】インターネットバンキングで振込処理を行う例における対話処理の流れ図。

【図17】インターネットショッピングで検索を行う例における対話処理の流れ図。

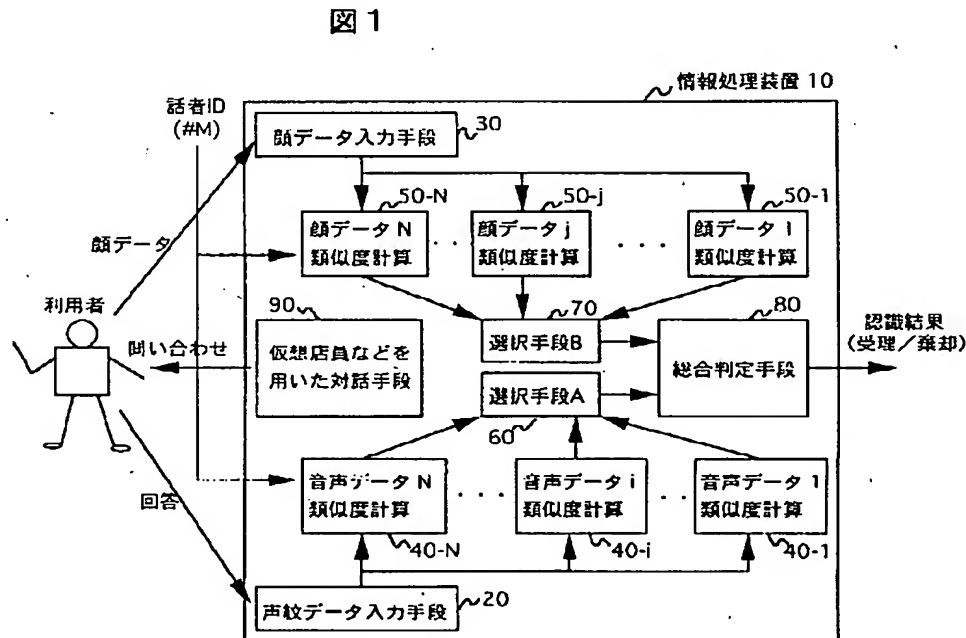
【図18】生体情報選択処理の一例を示す流れ図。

【図19】本発明の第2の実施例となる情報処理装置を示すブロック図。

【符号の説明】

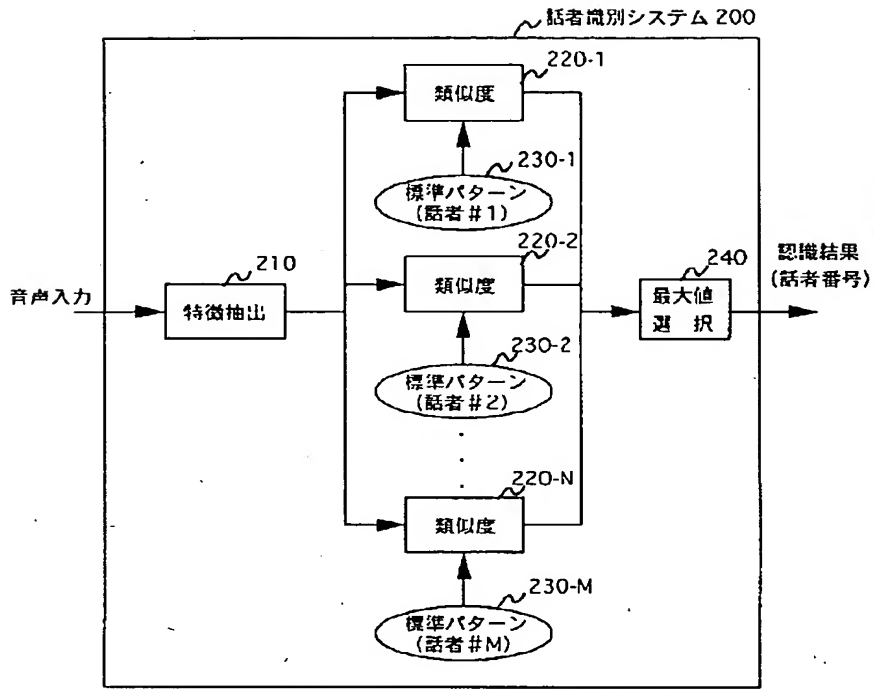
10…第1の実施例となる情報処理装置、20…生体情報データ入力手段、30…生体情報データ入力手段、40…類似度計算システム、50…類似度計算システム、60…類似度選択手段、70…類似度選択手段、80…総合判定手段、90…対話手段、110…特徴抽出部、120…類似度計算部、130…標準パターン、85…更新手段。

【図1】



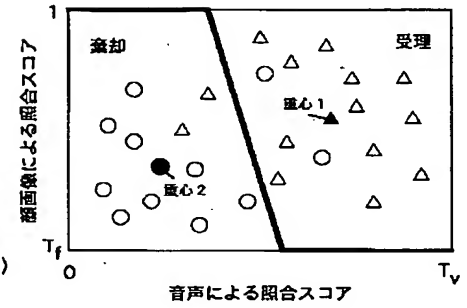
【図2】

図2



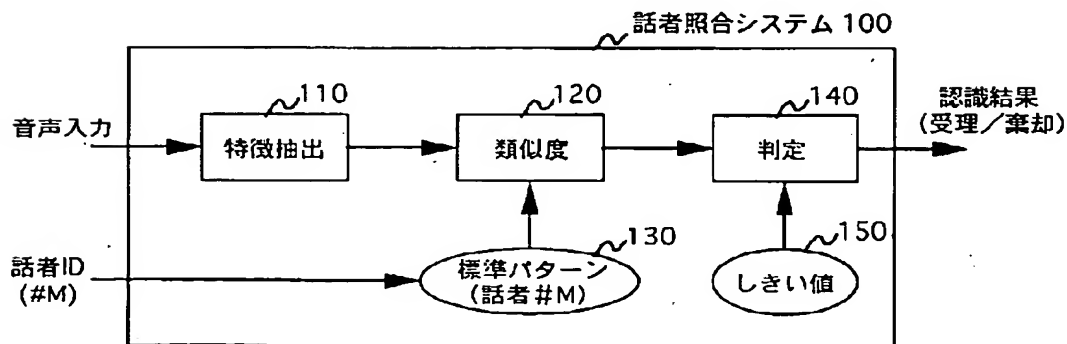
【図7】

図7



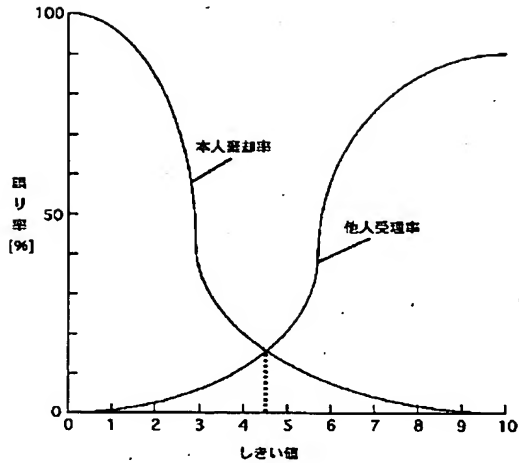
【図3】

図3



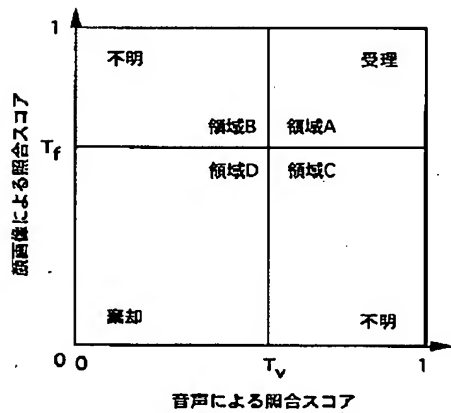
【図 4】

図 4



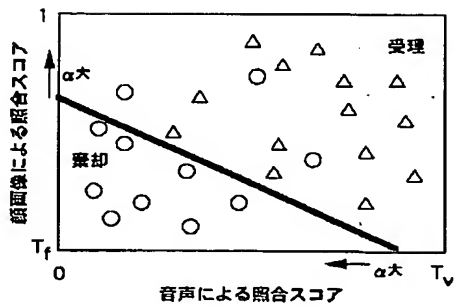
【図 6】

図 6



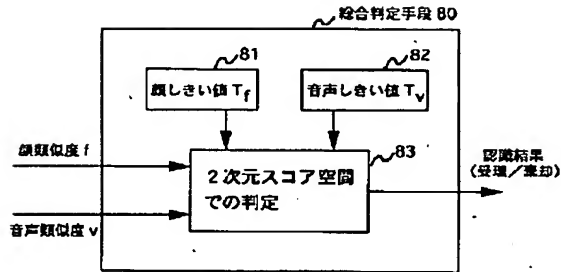
【図 11】

図 11



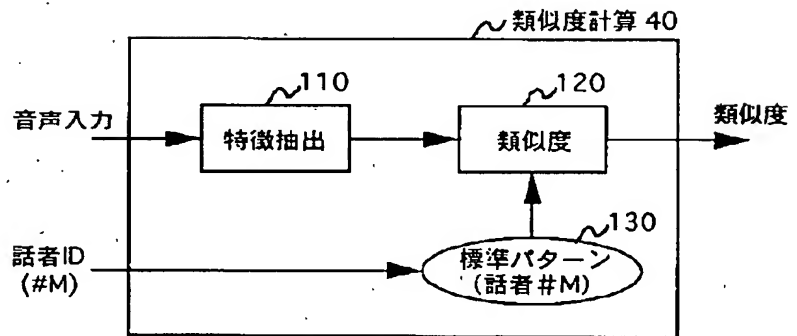
【図 5】

図 5



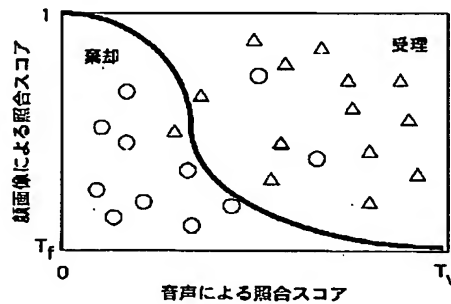
【図 8】

図 8



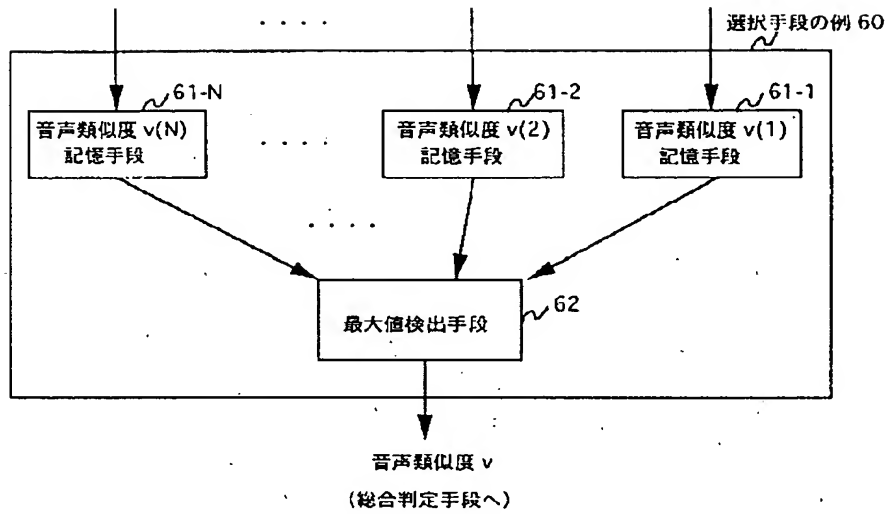
【図 13】

図 13



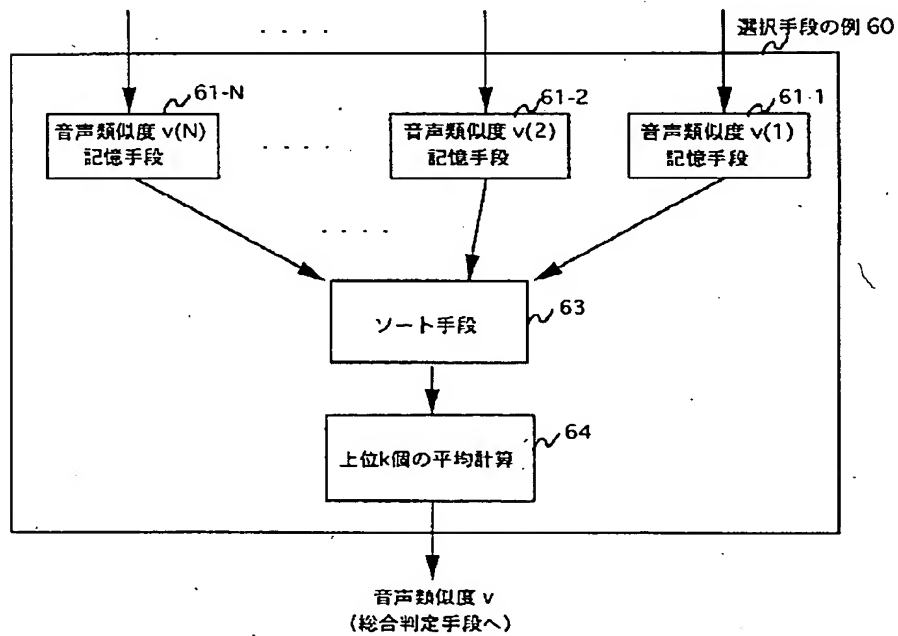
【図9】

図9



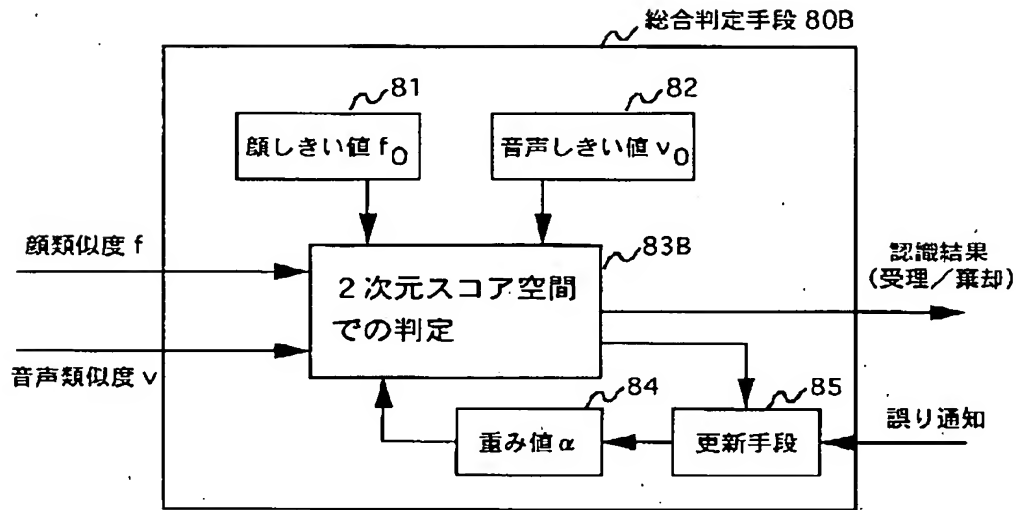
【図10】

図10



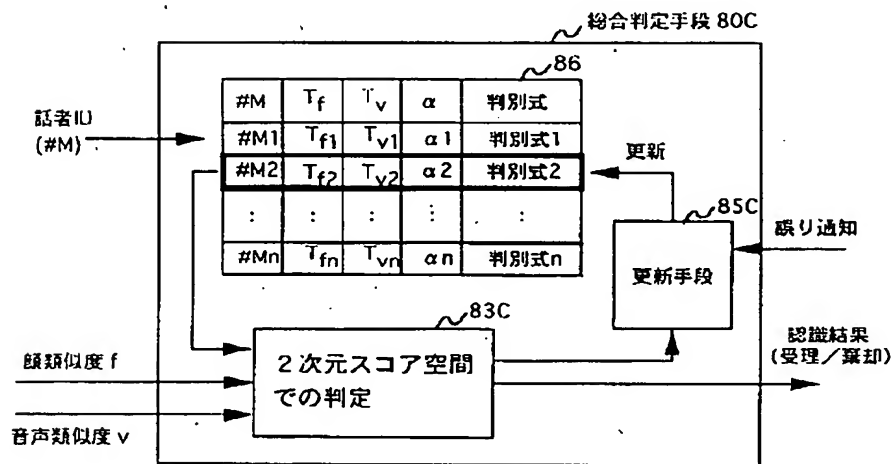
【図 12】

図 1 2



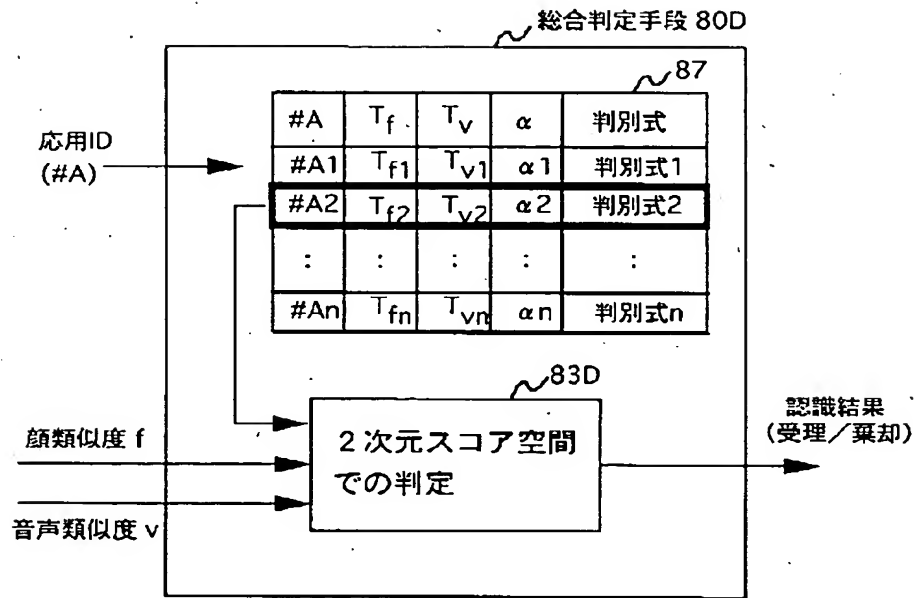
【図 1 4】

図 1 4



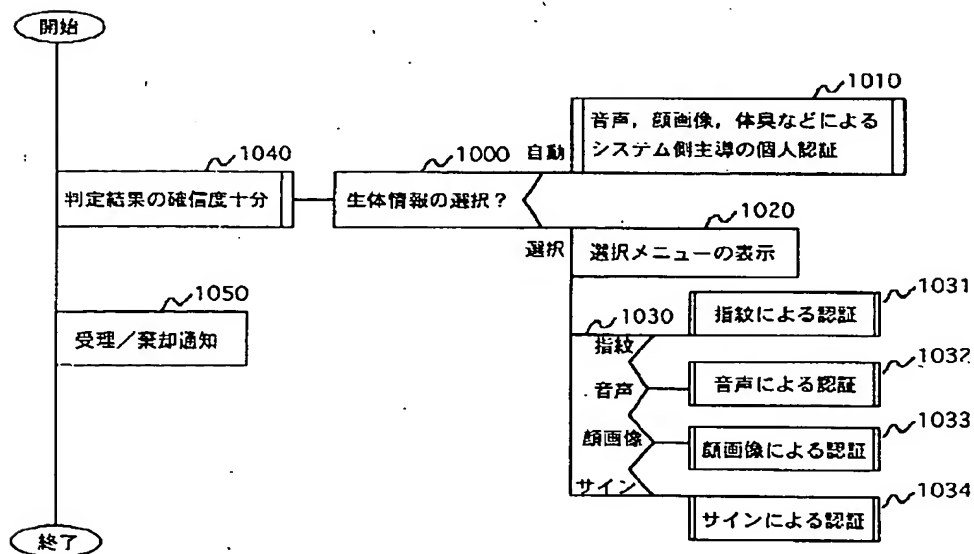
【図15】

図15

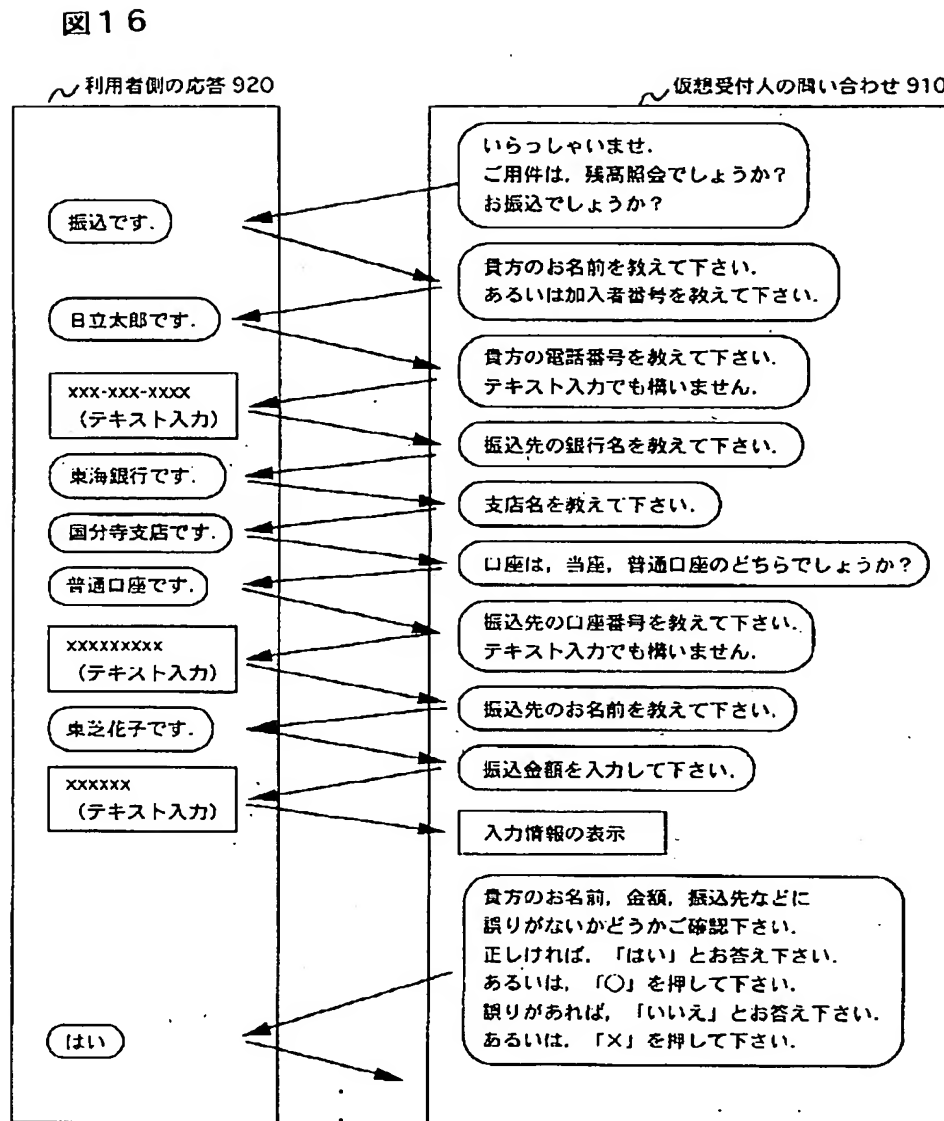


【図18】

図18

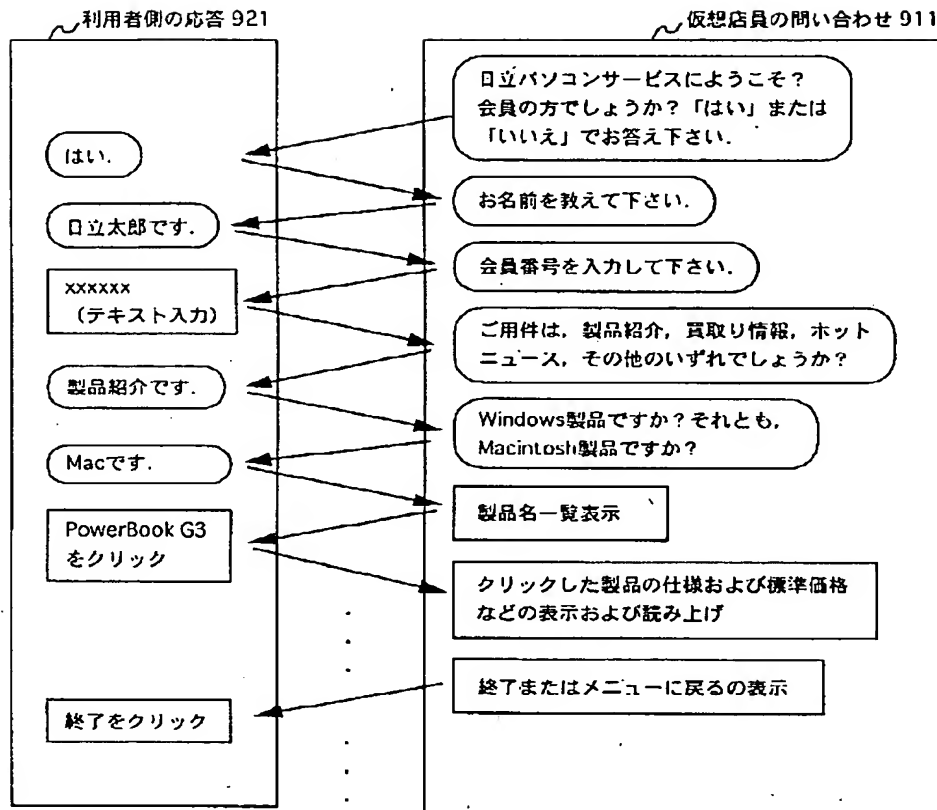


【図16】



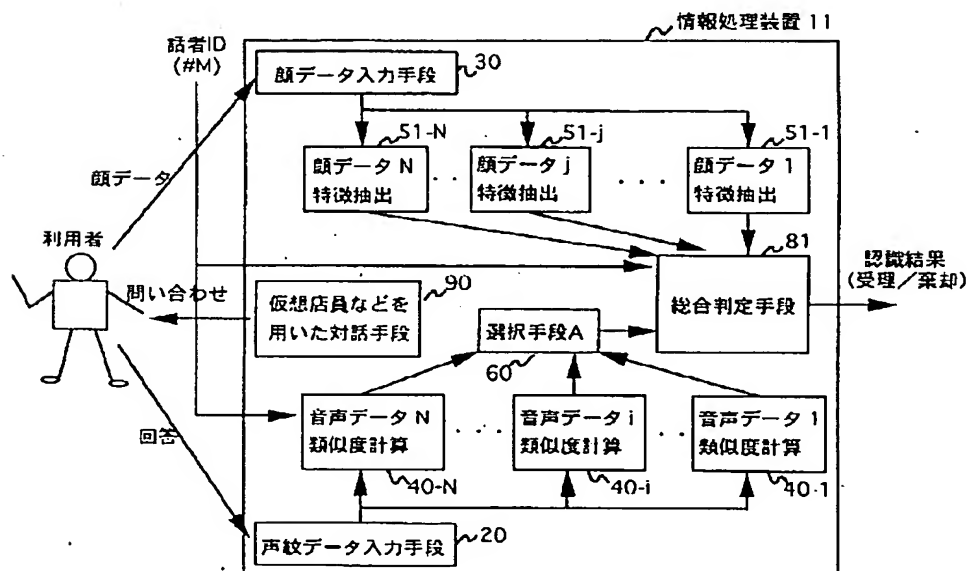
【図17】

図17



【図19】

図19



フロントページの続き

| (51) Int. Cl. ⁷ | 識別記号 | F I | テ-マ-ト (参考) |
|----------------------------|-------|---------------|-------------------|
| G 0 6 F 3/16 | 3 2 0 | G 0 6 F 15/00 | 3 3 0 F 5 E 5 0 1 |
| 15/00 | 3 3 0 | A 6 1 B 5/10 | 3 2 0 B 5 L 0 9 6 |
| G 0 6 T 7/00 | | G 0 6 F 15/62 | 4 6 5 A |
| G 1 0 L 17/00 | | 15/70 | 4 6 5 A |
| | | G 1 0 L 3/00 | 5 4 5 F |

(72) 発明者 炭野 重雄
 東京都国分寺市東恋ヶ窪一丁目280番地
 株式会社日立製作所中央研究所内

F タ-ム (参考) 4C038 VA07 VB03 VB04 VB05 VB40
 VC05 VC20
 5B043 AA09 BA01 BA02 BA03 BA04
 BA06 BA07 FA08 GA02
 5B047 AA23
 5B085 AC03 AE06 AE23 AE25
 5D015 AA03 DD02 DD05 GG01 HH05
 LL02
 5E501 AA09 AB22 AC33 AC42 BA12
 CB05 CB14 CB15 CC14 DA11
 DA15 DA17 EA21 EB01 FA32
 5L096 BA16 BA18 HA09 JA03 KA04

This Page Blank (uspto)